2014

Computing Trust as a Form of Presumptive Reasoning

Pierpaolo Dondio
Technological University Dublin, pierpaolo.dondio@tudublin.ie

Luca Longo
Technological University Dublin, luca.longo@tudublin.ie

Recommended Citation

Dondio, P. & Longo, L. (2014). Computing trust as a from of presumptive reasoning. *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*11-14 August, Warsaw, Poland. doi:10.1109/WI-IAT.2014.108

Computing Trust as a form of Presumptive Reasoning

Pierpaolo Dondio
School of Computing
Dublin Institute of Technology, Dublin, Ireland
pierpaolo.dondio@dit.ie

Luca Longo
School of Computing
Dublin Institute of Technology, Dublin, Ireland
luca.longo@dit.ie

Abstract— **This study describes and evaluates a novel trust model for a range of collaborative applications. The model assumes that humans routinely chose to trust their peers by relying on few recurrent presumptions, which are domain independent and that form a recognisable *trust expertise*. We refer to these presumptions as *trust schemes*, the specialised version of Walton's *argumentation schemes*. Experimental evidence is provided about trust schemes efficacy with a detailed experiment over an online community of 80.000 members. Results show how proposed trust schemes are more effective in trust computation when they are combined together and when their plausibility in the selected context is considered.**

Keywords—computational trust, online communities, fuzzy logics

I. INTRODUCTION

Computational Models of Trust have emerged in the last decade with the aim of exploiting the human notion of trust into open digital worlds. *Trust*, as intended by the computational trust community, is a *prediction* that the trustee entity will fulfill the expectations of a trustier in the context of a specific task.

A *trust computation* quantifies the level of trustworthiness of a digital entity, called a *trust value*. This computation requires the identification of the appropriate input data, the *trust evidence*. These data are in general domain specific and the result of an analysis conducted over the application involved. The selection of evidence and the subsequent trust computation are informed by a computational *trust model*.

This paper describes a novel trust model applicable to a range of Web applications. The main idea is the assumption that humans trust their peers by relying on few recurrent presumptions which are domain independent and that forms a recognizable *trust expertise*. We refer to these presumptions as *trust schemes*, the specialised version of *argumentation schemes*, notion proposed by Walton [19]. Example of trust schemes applicable to virtual identities are: *reputation*, *past-outcomes*, *degree of activity*, *degree of connectivity*, *regularity*, *stability and accountability*. The goal is to provide experimental evidence to answer the following research question: *"are trust schemes effective in computing trust?"*

Answering this question requires (1) defining a meaningful list of trust schemes, (2) showing a way to compute them, and (3) experimentally evaluating them. This work addresses these three issues: it provides a list of trust schemes that, although not exhaustive, is adequate to support meaningful trust metrics; it provides a framework to compute such schemes based on the notion of critical questions and fuzzy inference, and it provide a detailed experimental analysis based on a large online community. In particular, our evaluation shows how a small subset of easily computable metrics, such as *Persistency* and *Activity*, are an accurate proxy for a multi-faced concept such as Trust. Our experimental evidence could help social scientists understanding key factors impacting the perceived trustworthiness of virtual identities.
Our solution is a knowledge-based system, and its success depends on the ability to match trust schemes to the application domain under investigation. The use of trust schemes help to decouple the above matching problem by requiring expertise only in the application domain and not in the context of trust computations.

Moreover, the instantiation of trust schemes show recurrent pattern across a large class of web 2.0 applications. For instance, the results of this paper are generic enough to be applicable to any forum-like online application.

The remaining of the paper is organised as follows. Section 2 describes the starting assumptions on the notion of trust followed by, in section 3, a list of trust presumptions believed to be useful for trust assessment. Section 4 describes the central notions behind trust schemes while section 5 is aimed at designing a computational framework for trust computation based upon trust schemes. Section 6 describe experiments and the evaluation of the proposed solution followed by a description of related works in section 7. A conclusion summarises the paper and highlight future works.

II. ASSUMPTIONS ON THE NOTION OF TRUST

One of the most comprehensive definitions of trust is found in Romano [15]. According to Romano, *trust is a subjective assessment of trustee's influence about the significance of trustee's impact over trustier's (potential) outcomes in a given situation, such that trustier's expectation and inclination toward such influence provide a sense of control over the potential outcomes of the situation*. The definition stresses the notion of trust as a complex evaluation involving trustee, trustier and context. Compatible with this definition, we made four basic assumptions underlying our trust system:

1) Assessing Trust is a reasoning process

Saying that trust is a form of reasoning seems to clash with intuition. Often humans take trust decision without reasoning, following a instinct, intuition, unconscious actions as described by Lagerspetz [13]. Anyway, when it comes to rational agents, trust must be a rational decision grounded on evidence.

2) Trust is a form of defeasible reasoning

Trust is a form of defeasible reasoning because it is made up of assertions that are presumptions not deductively valid, but whose validity can be attacked or supported by new evidence. Trust computation could therefore benefit from studies in defeasible argumentation, such as Walton [19].

3) Trust is a distinct expertise with proper patterns

Trust is a distinct form of knowledge per se, an expertise that humans adopted in their decisions. As a form of knowledge, it can be modeled by recognizing recurrent patterns,

mechanisms and rules. This third assumption places this paper in the line of work commenced by Marsh [14], Castelfranchi and Falcone [11], where trust is a cognitive human *phenomenum* with proper ingredients and rules

4) Trust can be approximated analysing footprints left by entities in a certain domain

We presume that entities leave footprints in the domain they interact that are enough to perform a trust assessment.

III. The Ingredients of Trust

Previous section stresses how trust is an expertise made of recurrent presumptions. This section provides a list of such presumptions useful to assess trust. This list does not aim to be comprehensive, but large enough to support a meaningful trust computation. We refer to these presumptions as *trust schemes* to maintain the analogy with the notion of argumentation scheme found in argumentation theory [19]. Table 1 shows a list of schemes categorised in different areas.

TABLE I. Trust Schemes

Time-Based	
Trust Scheme	Trustee's presumption
Longevity	Trust entities with high longevity
Persistency	Trust entities acting persistently
Regularity	Trust entities acting regularly
Stability	Trust stable entities
Information-Sharing	
Indirect Experience	Trust entities according to other's people recommendations
Reputation	Trust entities with high reputation
Social-Role	
Authority	Trust entities with high authority
Connectivity	Trust entities that are well-connected in the environment
Popularity	Trust popular entities
Visibility/ Accessibility	Trust entities that are visible and easily accessible
Transitivity	Trust what your trusted entities trust
Information Provisioning	Trust entities that provide /share information
Activity-Based	
Pluralism	Trust entities or objects that are the results of many points of view
Activity	Trust active entities
Pertinence	Trust entities whose activity is pertinent to the domain
Outcome-based	
Past-Outcomes	Trust entities that did well in the past
Prejudge- and Grouping-based	
Similarity	Trust entities similar to the trustee
Categorization	Trust an entity on the base of the category it belongs to
Standard Compliance	Trust an entity that satisfies a standard
Similarity to Trust	Trust what it is similar to what the trustee trusted
Game-Theoretical	
Common Goal, Risk or Situation	Trust an entity that shares similar goals, risks or situations
Cost/Benefits	Trust an entity if it has a favourable benefit/cost ratio for the situation
Fulfillment	Trust entities that are committed to fulfil the task assigned
Risk Profile	Trust entities with a compatible risk profile

Time-based trust schemes

Trust is a question of time. This class of schemes builds trust arguments using only information about time, usually temporal intervals between interactions or interactions' timestamp. They do not consider *what* was done during an interaction and – more importantly – *how* it has been done. The focus is on *when* it happened.

The time-based trust schemes are *longevity*, *regularity*, *persistency*, *stability*. The importance of time-based information for assessing trust has been acknowledge by Carter [9], Longo [8] and by the common sense. The schemes augment the perceived accountability and experience of the trustee generating a positive argument to trust.

Trust schemes based on information sharing

This class encompasses the classical recommendation and reputation systems and all the solutions based on third-party information. Trust is derived by the *indirect experience* of trustworthy third-parties (see [1] for an up-to-date review).

Trust schemes linked to social role

Schemes in this class suggest that a trustee should be not judged in isolation but for the links and roles he/she has in the environment he/she is interacting in. Others entities may guarantee for him/her, or its public role may give assurance that the entity is *for real*. The core evidence we believe should be collected is: trustee's acquaintance, to whom it is linked and interacts, if it has specific roles in the environment, how easy it is to access and contact the entity and how transparent the information he provided is. In the current landscape of trust models, the *sociogram* of Sabater [10], the approaches based on network analysis and some trust factor proposed by Carter [9] strongly informs the definition of this class of schemes. As Carter [9] wrote *"the reputation of an agent is based on the degree of fulfillment of roles ascribed to it by the society"*. The trust schemes proposed in this section are: *authority*, *connectivity*, *popularity*, *accessibility/visibility* and *transitivity*. Their computation may rely on network analysis metrics such as various centrality measures as employed by Golbeck [3].

Trust schemes based on activity analysis

This group of trust schemes focuses on the activity of each entity in the environment, i.e. *what* an entity did rather than *when* or *how*. It focuses mainly on quantitative aspects, not considering the outcomes of an action but rather the quantification of the activity of an entity in the environment. Trust schemes proposed in this area are: *pluralism* and *activity*. The former refers to whether the information produced is the results of many opinions or actions. The latter is a clear ingredient of trust: it increases accountability, experience, familiarity with the environment.

Trust schemes based on (past) outcomes

This class contains the classical past-outcomes trust predictions. The scheme is usually implemented by using Bayesian models to update trust beliefs in the light of new interaction outcomes. Recently Dampster-Schafter models have also been investigated (see [1] for an up-to-date review).

Trust schemes based on statistics and grouping

These set of trust schemes ground their assumptions on the statistical significance of some properties of the trustee compared to other entities or group of entities. The sociological motivation behind this class is the socio-psychological studies of Kahneman and Tversky [13], the use of *categorization* in Castelfranchi and Falcone [12] and the concept of *prejudice* in computational trust as used by Sabater [10]. Entities trust other entities on the basis of the categories they belong to, or on the basis of similarities/dissimilarities with the trustier entity. The common idea behind these mechanisms is that trust can be transferred among similar entities/situations and properties can be assigned to an individual based on signs that identify that individual as a member of a given group [16]. This class of trust schemes encompasses *Similarity*, *Similarity to Trust*, *Categorization* and the *Standard compliance* trust scheme. They are all based the concept of similarity quantification. *Similarity* analyses the similarity between the trustee and the trustier, therefore it reflects a local point of view. *Similarity to trust* analyses the similarity between the trustee and a stereotype of the trustworthy entity that the trustier build in its mind. *Categorization* assesses the similarity between the trustier and a group of entities. Finally the *Standard compliance* trust scheme assesses the similarity between the trustee and an *accepted* standard present in the environment.

Trust schemes based on Game theory and Cognitive models

The trust schemes in this class consider opportunistic motivations that the trustier and the trustee may have in a situation, modeled as a game among rational players. The assumptions behind these trust schemes is that the trustee and the trustier are both rational entities that are trying to maximize their satisfaction and minimizing the effort spend. Therefore, the understanding the cost and benefit of the other entities produces an argument in favour or against trust.

IV. The Structure of Trust Schemes

Argumentation schemes were described, among the others, most notably by Walton [19]. Walton defines argumentation schemes in the context of his analysis of presumptive reasoning of which a trust-based decision is an instance. He notes how presumptions are rarely ad-hoc constructs that are used in a dialogue. More often, presumptions are instances of generic patterns of reasoning defined as the glue that holds argumentation together and makes it reasonable. Examples of his argument schemes include *argument from popularity*, *expert opinion*, *ad ignorantiam*. A set of critical questions tests the assumptions on which a scheme bases its plausibility. They are inherent to the argumentation scheme and their role is to rebut or make the argument generated by each scheme stronger. Both critical questions and argumentation schemes have to be matched to some evidence/fact of the domain.

The trust schemes proposed in this study are a specialized version of argumentation schemes. They can be seen as defeasible rules supporting either trust or distrust of an entity. They are indeed defeasible, since they have exceptions and they are based on assumptions. For instance "*I trust this baker shop since it has been always full of customer*" is an instance of the popularity trust scheme. It is a defeasible conclusion whose plausibility varies based on the context. Yet "*the shop is the only one in town"* or "*the shop next door is empty*" respectively decrease or increase the scheme plausibility. Therefore the strength of a scheme conclusion is proportional to the strength of the evidence used (how full is the shop) and the plausibility of the scheme in the context (is popularity a sign of trust here?).

The above observation suggests implementing a scheme-based computation into a three-stage process. In the first stage, each trust scheme, representing a defeasible rule, is matched over the available elements of the application domain. This stage is referred to as evidence selection. An element can be instantiated by more than one trust scheme and vice-versa. Elements of the domain could be directly used in a trust scheme or more complex intermediate computations can be performed to match the scheme. In the second stage, the identified trust schemes are tested against their critical questions to estimate their plausibility. This stage may require information coming from the application. In the third stage, the tested schemes are aggregated into a final trust value.

As an example of trust scheme, we consider the *past performance* trust scheme, the most used in literature and regarded as the *most objective*.

Defeasible Presumption. Entities that did well in past interactions will (*presumably*) do so in the future, since they showed the ability to fulfill expectations.

Computation: how to quantify it? In computational trust literature, the scheme is usually (but not exclusively) implemented by counting good interactions (p) and bad past interactions (n). The value of trust is usually represented with a beta distribution whose two characteristics values are n and p.

Critical Questions. Each trust scheme has a set of critical questions aimed at testing its validity. The *past- performance* scheme is indeed a presumption. Its critical questions include checking whether the interactions are out of date; if they are relevant to the current context; if the trustee has somehow changed; if the trustee is motivated; if external constraints outside of trustee's control affected its past performance; the difficulty of each past interaction. It is important to note how the investigation, started by the critical questions, suggests also ways to improve a scheme computation.

TABLE II. Critical Questions from other trust schemes

Trust Scheme	Critical Questions	Description
Longevity	Is x active?	In absence of activity, longevity is not an evidence for trust
Stability	Is x active?	In absence of activity, x can be stable since out of business
Stability	Is x persistent?	As above
Past performance	Is x Persistent? Is x Active? Is x Stable? Is x Pertinent?	Past performance are a weaker evidence if the entity changed, is not very active and persistent
Reputation	Is x Persistent? Is x Active? Is x Stable?	As above
Reputation	has x good past-performance?	Direct experience is usually regarded superior than indirect experience

Trust schemes are not isolated rules, but rather there is strong mutual dependency among them. One of our hypotheses is that a trust assessment is stronger if the relations among trust schemes are taken into account. These mutual relations are no

more than additional critical questions. For instance, the trust scheme *past performance* mentioned above is affected by the value of the trust scheme *Stability*. Table 2 presents a list of critical questions among trust schemes.

V. COMPUTING SCHEMES

This section presents how to compute a *trust value* using trust schemes equipped with critical questions. The proposal follows the recent work of Prakken on the nature and representation of argumentation schemes [20].

Each trust scheme represent a defeasible modus ponens rule of the kind $(A, A \rightarrow T) \rightarrow T$. The conclusion T means *trust entity x* and it can be replaced by $\neg T$ (*distrust entity x*). A is the premise of the scheme, based on evidence collected in the context under consideration. The second premise $A \rightarrow T$ contains the defeasible assumption encoded in the scheme, that links a piece of evidence used in A to the conclusion *trust x*.

Let us provide an example using again the *past performance* mechanism. The scheme is: A - "*Mark has high past performanc*e" and $A \rightarrow T$: "*high past performance implies trust"* and therefore we conclude that Mark deserve our trust. The implication $A \rightarrow T$ is clearly a presumption not valid in presence of other pieces of evidence (such as "*all the past performances refer to an irrelevant context*"). These pieces of evidence are exactly the critical questions, which therefore result as evidence invalidating the trust scheme assumption. More precisely, following [20], a scheme of the kind $(A, A \rightarrow T) \rightarrow T$ can be attacked in the following ways:

1. by undercutting the reasoning link $A \rightarrow T$, that means by finding exceptions or situations in which that assumption is not valid. An undercutting attack leaves unchanged the premise A, but it just invalidates the reasoning link.
2. by contradicting the conclusions T, for instance using another argument that suggests $\neg T$.

It is important to note how, in our settings, the scheme cannot be attacked by stating $\neg A$, since we assume A to be a verified fact (not an assumption) based on evidence from the context domain where a trust metrics has to be computed.

Trust scheme as fuzzy inference rules

If we look at the above scheme, it is obvious how terms involved are indeed vague and experienced at different degree. For instance, an entity is *active*, *stable* or *reputable* to a degree. The plausibility of the assumption $A \rightarrow T$ encoded in each trust scheme is also perceived at different degrees of plausibility. We therefore propose to treat trust scheme as a fuzzy inference rule. A fuzzy variable, such as *height*, *weight*, is a quantity that can take linguistic terms, such as *high*, *low*, *medium*. Each fuzzy term is described by a fuzzy set. A fuzzy set is a pair (U, m) where U is a set and $m: U \rightarrow [0,1]$ is the membership function that assigns to each element of $x \in U$ a degree of truth $m(x)$, quantifying to which degree x is an element of the fuzzy set. U is called the universe of discourse.

Each of the evidence used in our model – serving as premises for trust schemes or their CQs – are fuzzy variables, such as *activity*, *reputation*, *stability*, *validity*, *trust/distrust* and they can take the linguistic values *high*, *medium*, *low*. The universe of discourse U of each linguistic variable depends on its domain. For instance, the universe of discourse of the term *activity* in the context of an online Web forum could be the number of messages posted by a user. Figure 1 shows the membership functions for the terms *low*, *medium* and *high* (for simplicity we work with triangular functions). The membership functions return the degree of truth of each element of U. For instance, a user x with 1000 messages could be perceived to be *highly* active with a degree of 0.8, while user y with 700 messages is *high* to a lower degree 0.5 and it is also a *medium* active user to a degree 0.2.

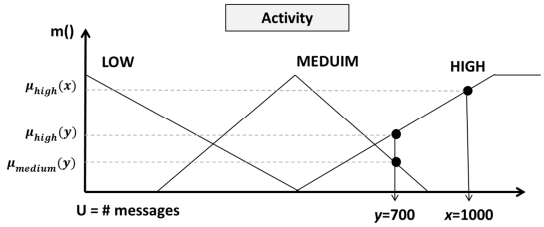

Figure 1 – The fuzzy variable activity and its terms

The universe of discourse of the variable $A \rightarrow T$ (the trust scheme assumption) is an index of plausibility in [0,1] derived from the critical questions analysis described later in this section. Finally, *trust/distrust* are also fuzzy variables and their universe of discourse is the interval [0..1], referred to as a trust level. Employing this terminology a trust scheme can be expresses as in the following form:

If A*: reputation of Mark is high and* $(A- > T)$*: the validity of reputation is medium then* T*: trust for Mark is medium*

In order to work with fuzzy inference systems, we have to quantify the degree of truth μ_A and $\mu_{A \rightarrow T}$ of the two premises A and $A \rightarrow T$. The quantification of such premises require an investigation of the application context where trust has to be computed.

For instance, a degree of *activity* of a user in an online forum application is quantified considering number of posts, discussions opened, attachments and so forth. The goal here is to quantify only the activity level, not trust. This task requires knowledge of the application domain only, while the trust schemes are aimed at computing trust. Regarding the level of plausibility of each scheme, a value is set according to how well the critical questions are answered. Each critical question is given a score on a *Likert* scale from 1 to 5, and subsequently the results of all the answered CQs are aggregated. Since each CQ is a reason that can undermine a trust scheme validity, even a single fully satisfied critical question can alone invalidate the scheme. Therefore, CQ do not accrue and the CQ with the highest value is considered. The degree of plausibility of a scheme T_{S_j} is therefore obtained by:

$$P\left(T_{S_j}\right) = 1 - \max_i \left(\frac{CQ_i^{T_{S_j}}}{5}\right) \qquad (1)$$

where we scaled the score $CQ_i^{T_{S_j}}$ of the i^{th} critical questions for trust scheme T_{S_j}. In case none of the CQ_i can be answered – not enough evidence available – a default value of 0.5 corresponding to a medium plausibility is used.

Computing a trust value using the Mandami Inference

Once it is known how to compute μ_A and $\mu_{A \rightarrow T}$ for all the trust schemes applicable and for all the trustee entities, we propose to use the Mandami inference system to derive a defuzzified value for the conclusions T and $\neg T$. In figure 2 an example of trust computation in the context of an online auction website is

depicted. In the example, three rules (representing three trust schemes) have been found to be applicable to a generic seller *x*. The rules conflict: two rules suggest *trust* x and one rule suggests *distrust* x. *Activity* (rule 1) and *reputation* (rule 2) are positive evidence, while x' low *past performance* (rule 3) is a negative evidence. The universe of discourse for *activity* is the number of items sold by x, for *reputation* is a reputation score found in the forum and for *past performance* we use a percentage of positive feedback received by x.

In order to compute a final trust value from these set of rules we follow the Mandami inference system (a comprehensive description can be found here [19]). The Mandami inference (in figure 2) uses *max* as conjunction (more precisely as T-norm operator) and *min* as disjunction (T-conorm) operators to combine fuzzy terms and rules.

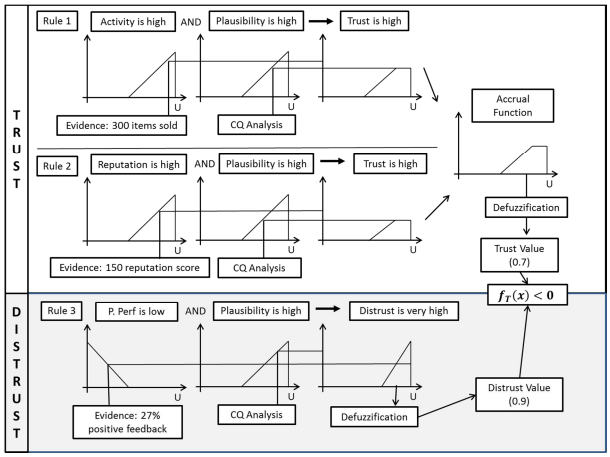

Figure 2 - Fuzzy inference to compute trust

In our context, for each rule and each entity i, the degree of truth of the conjunction of the two premises A and A$->$ T has total degree $h_i = \min(\mu_{A_i}, \mu_{A_i \to T})$. This value is used as an upper limit for μ_T, the membership function associated to the conclusion, that results $\mu_{T_i} = min(\mu_T, h_i)$. The procedure is repeated for the other rules supporting trust resulting in n membership functions. These *n* membership functions are then aggregated with the T-conorm operator to produce the final aggregated membership function $\mu_{T_{global}} = \max(\mu_{A_0}, \dots, \mu_{A_n})$, that is *defuzzyfied* to generate a trust value in its universe of discourse [0..1]. There is a set of popular defuzzyfication method in literature. In our evaluation we use the *mean of maxima* [19] method, that is the average of all the values d in U twhere $\mu_{T_{global}}(d)$ has a global maximum.

The procedure described above is repeated for the trust schemes supporting distrust as well, but the two set of rules (supporting trust or distrust) are kept divided and they are accrued separately. In fact, two schemes concluding *trust is high* and *trust is medium* both support the same fuzzy term and they do accrue (as they do arguments whose conclusions contains both the fuzzy variable *distrust*), while *trust is high* conflicts with *distrust is medium* and they require a different treatment.

We call $T(x)$ and $\overline{T}(x)$ the final defuzzified values for trust and distrust for entity x. $T(x)$ quantifies the reasons to trust an entity while $\overline{T}(x)$ the reasons not to trust it. A final decision is then made comparing the two values. A skeptical trustee would require $T(x)$ to be high but also $\overline{T}(x)$ to be null (or below a threshold), while a credulous trustier will only look at $T(x)$. We introduce a trust evaluation function f_T to join the values of $T(x)$ and $\overline{T}(x)$, useful to compare two trustee entities. The function requires that both the difference between $T(x)$ and $\overline{T}(x)$ be high (representing low conflicts) and the value of $\overline{T}(x)$ low. The final *function of trust evaluation* f_T proposed is:

$$f_T(x) = \big(1 - \overline{T}(x)\big)\big(T(x) - \overline{T}(x)\big) \qquad (2)$$

VI. EVALUATION

We evaluated the efficacy of our trust model over the large online community *FinanzaOnline.it*, with a dataset of about 80.000 registered users and about 9 million messages. Aim of the experiment is the computation of a level of trustworthiness for each forum member. We quantify the efficacy of our model against an explicit poll, asking forum members to identify trustworthy entities. The anonymous poll received almost 1.500 answers from 298 users. The results of the poll showed a clear consensus about the most trustworthy entities. According to the votes received, we divided users in ordered tiers. The first tier contains the 10 most trustworthiness entities, the second contains the members from 11 to 50 positions. A trust computation is successful if it recognizes tier 1 and tier 2 members as the most trustworthy. We evaluate the accuracy of our metric using the following mean squared error metric:

$$E(n) = \frac{1}{n}\sqrt{\sum_{x=1}^{n} C_{rank}(x) - T_{rank}(x)} \qquad (3)$$

where n is the number of members included in the metric, $C_{rank}(x)$ is the rank of member x according to the community survey, $T_{rank}(x)$ is the rank according to our trust computation. Therefore *E(n)* measures the average error generated considering the set of top-*n* members only.

Trust Schemes Engineering

FinanzaOnline is a typical online forum where users can post, attach, open polls and have a public profile. The forum is divided into a stock market-related zone and a *free chat* zone.

TABLE III. AVAILABLE APPLICATION ELEMENTS

$N_p, N_{att}, N_{poll}, N_{3D}$	Number of posts, attachments, poll opended and threads started
$N_{free}\ N_{trading}$	Number of messages in the free-zone, number of trading messages
t_{last}, t_0, t_{now}	Time of last posts, time of registration, present time
$\bar{L}_p$	Average length of posts
$A_{news}, A_{graph}, A_{other}$	Number of attachments containing news, graphs or other
$Ce(u), Cl(u), Ib(u)$	Centrality, closeness and in-between centrality of user u in the network build using users citations
$A_{skype}, A_{pic}, A_{bio}$	the presence contacts information (email, skype), pictures, biographical information
r_a	Reputation level of user a

Using the available application elements, we matched and engineered a set of trust schemes and we assigned a plausibility value. Table 3 presents the set of evidence used by our trust schemes, deducted by the underlying set of domain elements present in the fourm. The majority of the evidence used, except

A_{graph}, A_{news}, has direct mapping with element available in *Finanzaonline.it*. A_{news}, A_{graph} were manually sampled to discriminate between attachments not related to finance, news or graphs. $N_{trading}$ is the number of messages related to finance written in the stock market section of the forum.

We adopted a percentile-rank method to quantify the strength of the evidence collected. This means that we rank users by each piece of evidence selected (for instance number of messages) and we consider a percentile score in [0,1] for each user. The percentile score is also used as the universe of discourse for all the schemes.

Longevity

Longevity is the interval between the time of last post and the time of first post. The plausibility, as it emerges from critical questions analysis is high, since the environment is selective; with a decreasing population of users during period of stock market crisis.

Persistency/Regularity

The scheme divides the timeline into intervals of equal size, equal to 1 day, 1 week or 1 month and computes the percentage of intervals in which the entity is active. As in every online community, *Persistency* is a strong argument and the critical questions analysis assigns to it a high level of plausibility. The presence of cycle of activity (5 days a week for instance) has been adopted in our experiment. The data available are complete and certain. The action chosen for detecting activity is the action of posting a message. Passive actions such as login are not considered. The same plausibility value is assigned to the complementary trust scheme *Regularity*. An entity is regular if the time interval between two consecutive interactions is relatively constant and not subject to high variance.

Activity

Activity was mapped considering the following indicators: posting a message (N_{post}), opening a discussion (N_{3D}), opening a poll (N_{poll}) and adding attachments to messages (N_{att}). The critical question analysis set the plausibility of the scheme high – in any online community contribution is seen as the cornerstone of trustworthiness, see [8.9]. Regarding the plausibility of the computation, the problem is to choose the appropriate *accrual function* for the 4 indicators of activity. Our analysis of the forum suggests that the action of posting message is the basic compulsory action (better computed also considering the size of the messages instead of the crude number). Entities that do not post messages cannot be considered active. The action of attaching a file to the post is optional; its value is only used to increase the strength of an entity but not decrease it. The action of opening a discussion/pool is an advanced action that is again optional, and therefore it is used as an positive evidence to strengthen certain entities.

Pertinence

Pertinence requires quantifying the extent to which the activity of a user is pertinent to the domain of *online trading* (theme of the selected forum). It does not try to understand whether an entity is a skilled trader, but rather whether he/she posts about trading and not something else. The scheme has high plausibility. We consider a user pertinent if:

1) it has a high number of trading messages or a low percentage of messages in the free-chat section, and
2) it has high number of news attachments and graphs, and a low number of non-trading attachments

Connectivity

The scheme relies on network metrics to quantify the prestige of users in *Finanzaonline*. We build a directed graph network where nodes represent members of the forum and a link from A to B means that user B cited a post p written by user A. Links are weighted by the number of times user A cites user B. Connectivity aggregates the rank of each users according to their *in-degree centrality* (measuring the number of quotes received by a user; self-citations are excluded), *in-beetwness centrality* (quantifying how the user is crucial in connecting different sub-group of users) and *closeness centrality* (measuring how close a member is to all the other members). The critical questions analysis sets the plausibility of the scheme to high.

Reputation

An internal reputation system is available. However, we do not use this information as trust evidence since our evaluation is already based on explicit user feedback and thus we have to avoid a circular argument. Moreover, our analysis of the internal reputation systems shows its lack of plausibility, revealed by the fact that the produced values are highly positively biased, and by the low acceptance of the system by the users. As a test, we include recommendation in our evaluation to study its effectiveness, that we expect poor.

Accessibility

Accessibility was mapped over the profile of each member. A Boolean score is given to the presence of 3 classes of evidence $A_{skype}, A_{pic}, A_{bio}$. These are considered all ingredients of the projections of a person into the online community, and they are seen by a strong majority of sociologists as fundamental aspects of trust. However, since the information is not verifiable, often malicious and incomplete, the overall plausibility of the scheme is poor and we expect better results by excluding it.

Results evaluation

We computed each trust scheme and a global trust value for each member of the community. The scope of our analysis is (1) to identify which trust schemes are more effective in assessing trust of online members, (2) to understand whether the consideration of the plausibility value of trust schemes has an effect on our results and, (3) to understand the impact of different aggregation strategies.

Table 4 and table 5 show the results of our experiments, globally (table 4), and for each trust scheme (table 5). The best case shows a value for *E(10)* of 3.4 using a set of 5015 users, meaning that the difference between our trust computation and users opinion is extremely narrow. Table 4 presents the overall results with or without the critical questions analysis, and with a credulous attitude (only positive evidence to support trust are considered) and skeptical (both negative and positive evidence are used and aggregated into the function f_t). The introduction

of critical questions makes the results more efficient. If we consider the computation without them – that means all trust schemes considered the same in terms of plausibility – the overall results have, in the best case, an average error of more than 80 positions for *E(10)* and more than 100 for *E(50)*. The main reason is the usage of two implausible trust schemes in the context, such as reputation and accessibility. Time-based and activity-based schemes were very effective individually.

TABLE IV. GLOBAL RESULTS

	Without CQs		**With CQs**	
	E(10)	*E(50)*	*E(10)*	*E(50)*
Credulous	112.8	171.4	4.1	65.3
Skeptical	88.7	143.2	**3.4**	39.7

TABLE V. RSME FOR EACH TRUST SCHEME

	Trust Scheme	**E(10)**	**E(50)**	**Pla**
Time-Based	Longevity	112	400	0.7
	Persistency	31	120	0.9
	Regularity	48	109	0.9
Activity-Based	Activity	19.6	91.8	0.7
	Competence	75	211	0.8
Social-Based	Authority	104	398	0.8
	Accessibility	2147	1863	0.7
Others	Past-Outcomes	N.A.	N.A.	0.9
	Recommendation	923	1803	0.2

Considering both positive and negative evidence sensibly improves the results. In table 3 *E(50)* is now reduced by 35% using a skeptical approach. A benefit is achieved for E(10), reduced drastically down to 3.4 positions. From a more detailed analysing of the results, it is possible to note how CQs give more consistency by

1. reducing the impact of entities with high but not regular activity (it is common to find entities that in one year wrote what normal entities write in 5-6 years and then they disappear), applied to 267 members;
2. by reducing the ranking of entities with high activity but low pertinence (applied to 197 members);
3. by excluding old but quite inactive entities, that still have good aggregated scores due to high longevity, pertinence or connectivity (applied to 1447 members).

Time-based Schemes. The 50 most trustworthy entities are all "old" ones. The forum of FinanzaOnline.it was opened in 1999, and the youngest of the top 50 entities registered in February 2004, while the average age is about 9.7 out of 12 years of forum life. 13 entities are more than 10 years old. Anyway, many other old entities are not trustworthy, so the scheme has only a one-way validity. Entities are persistent, the top 50 entities' average time of non-interaction is less than one week (5.3 days), and only three entities in the top 50 had an idle time longer than two months in their life.

Activity. The top 50 entities are very active. They hold the top five positions for the scheme. Also they usually - but not always - attach files to their forum messages. The top 50 users usually start conversations, and this makes the most significant difference with *ordinary* entities.

Competence. The top entities have good signs of competence. Anyway, 5 of the top 50 entities do not have a very high score. These entities show a good number of trading messages, but are also keen to chat and give contributions to other sections of the forum not related to trading. The community does not regard this as bad action, as far as they keep writing messages of high competence as well.

Connectivity. Surprisingly, the top 50 entities show a variable behavior in this factor. The top 10 entities perform well and are usually well quoted by a high number of members, but among the top 50, 5 of them have a very poor scoring. Despite of this, the community judged them among the more trustworthy. These entities have a good score in the other factors, but it seems they do not interact with other entities.

In conclusion, some of the schemes were effective in the computation, but high results were gained by combining them and by assessing its plausibility.

VII. RELATED WORKS

Computational Trust. The trust model proposed in this study is in line with the research of Marsh [14], Castelfranchi and Falcone [11] and the computational trust community. Our solution is an example of a non-reductionist approach to trust with some unique features. It is multidisciplinary embracing argumentation, fuzzy logics and, despite it makes use of previously tested techniques for trust computation, it also incorporates the notion of trust schemes and it computes and evaluates new trust mechanisms. In this respect, it investigates some trust scheme not previously adopted in the trust community such as temporal-based schemes, the use of pertinence, visibility, and connectivity. The proposed solution is also a meta-model that considers a broader view on trust. While reputation systems or past-outcomes analysis are indeed widely used mechanisms to compute trust, this work proposes complementary techniques. It is a framework that allows different pieces of evidence to coexist and it applies probabilistic and reputation-based approaches for the computation of trust. Its main features are:

3. the focus on the defeasible nature of each trust mechanism
4. the consideration of a method to check the plausibility of a mechanism in the selected context
5. the interaction of various techniques.

Our model falls in a category, non-necessarily in opposition to the probabilistic models, that includes few cognitive models that consider trust as a mental process with proper rules and content. These models provided a set of pieces of evidence and techniques that inspired the definition of our trust schemes and the design of our framework.

The cognitive trust model of Castelfranchi [11], for instance, considers trust as a distinct expertise composed by four basic beliefs: *competence*, *fulfillment*, *dependence* and *disposition*. These beliefs inspired the design of some trust schemes. Similarly, the work of Carter on information sharing communities [9] considers trust as an aggregation of five basic roles: social information provider, content provider, longevity role, administrative feedback role, interactivity role. Although these roles have a computational counterpart, they are not

systematically embedded into a trust computation. These influenced the design of our model as an extensible framework able to incorporate different roles into a trust computation. The *sociogram* by Sabater [10], based upon canonical reputation and past-outcome mechanisms, includes a set of new sociological pieces of evidence relevant to trust computation that inspires our trust schemes based on social roles.

Trust modeling as an argumentative process. In [23] the author was the first to propose the use of argumentation schemes in trust modeling. Here, a limited list of schemes was proposed and a preliminary evaluation on the Wikipedia project performed. In the last few years there has been a growing mutual attention between the argumentation and computational trust community. For example, the W4 EU-COST group [17] investigated agreement technologies, including trust as a key topic. A decision about trust is indeed an argumentative process, where conflicting pieces of evidence has to be reconciled. In this field we cite the work by Matt [18]. Stranders [5] and Villata [24] investigate the use of argumentation for trust computation from a formal point of view. The goal of these researches is the study of a theoretical argumentation model to suit the notion of trust, differing from our solution where trust metrics are computed. The only work that proposed (5 years after our proposal) a similar idea of trust schemes is [22]. Here authors pursuit the idea of using Walton-style argumentation schemes for trust analysis. Their work is experimental and the list of schemes and critical questions is rather descriptive to inform a computational model. Regarding the actual tools used in the trust computation, although fuzzy inference and the use of critical questions were adopted, these were separately applied and no effort was made to consider them into a unified framework [22]. Similarly, fuzzy sets have been used in trust representation, but fuzzy inference has not. Fuzzy logic has been also adopted in [4] to compute trust and applied over a real dataset. However, the proposed trust model is based on the past-performance mechanism where fuzzy sets are used to grab the uncertainty of input data and not in the inference process as we suggest in our work.

Finally, *t*he closest work at present, to the best of our knowledge, remains our previous research [26]. However, this previous study did not clarify how plausibility levels of each trust scheme were set, and it did rely on hard-coded plausibility values. It used simple algebraic operators to aggregate trust schemes. The solution presented in this paper completes [26]: it corrects its major flaws, changing and extending its computational abilities and it provides a new experimental evaluation of the computational model.

VIII. Conclusions and Future Works

In this paper we presented a knowledge-based system to compute trustworthiness of digital entities. Starting from a set of presumptions humans routinely use for assessing trust, we describe a model to deploy a trust metric around those presumptions, called trust schemes, in a target application domain. We provided an implementation of the model and reported about experimental evidence collected to date, showing how trust schemes could efficiently approximate the human judgment about trust in the context of a large online Web community. Our computation is application-contained and non-invasive, since it uses only domain elements, scrutable and able to suits various Web 2.0 application such as Wikis and Online fora. The method extends the trust computation in several ways. It introduces a broader set of evidence, it represents by novel trust schemes, along with the definition of the mutual relationships among the trust schemes. Future works will be in the direction of collecting a larger set of pieces of evidence and case studies to further understand the strengths and weaknesses of our model.

References

[1] Pinyol, Isaac, and Jordi Sabater-Mir. "Computational trust and reputation models for open multi-agent systems: a review." Artificial Intelligence Review 40.1 (2013): 1-25.

[2] P. Sztompka. Trust: a Sociological Theory. Cambridge University Press. Cambridge, UK, 2000

[3] Golbeck, Jennifer Ann. "Computing and applying trust in web-based social networks." (2005).

[4] Griffiths, Nathan. "A fuzzy approach to reasoning with trust, distrust and insufficient trust." Cooperative Information Agents X. Springer Berlin Heidelberg, 2006. 360-374.

[5] Stranders et al. Fuzzy Argumentation for Trust. Computational Logic: 8th International Workshop, pages 214-230, Springer-Verlag, Berlin, 2008 isbn 978-3-540-88832-1

[6] P. Dondio, "Trust as a Form of Defeasible Reasoning", PHD thesis, Computer Science Department, Trinity College, Dublin, 2009

[7] www.finanzaonline.it Last accessed on the 12 August 2008.

[8] L. Longo, et al., Temporal Factors to evaluate trustworthiness of virtual identities, IEEE SECOVAL 2007, part of SECURECOM 2007, Nice, France, September 2007

[9] J. Carter, E. Bitting, A. Ghorbani. Reputation Formalization for an Information-Sharing Multi-Agent System. Computational Intelligence 18(2), pagg. 515-534. 2002

[10] J. Sabater, C. Sierra. REGRET: A reputation model for gregarious societies. 4th Workshop on Fraud and Trust in Agent Societies, Montreal, Canada. pp. 61-69, 2001

[11] C. Castelfranchi, R. Falcone. Trust is much more than subjective probability: 32nd HICSS. Hawaii, 2000.

[12] P. Tversky, D. Slovic. Judgment Under Uncertainty: Heuristics and Biases. Cambridge University Press, UK,1982

[13] O. Lagerspetz, O. The Tacit Demand. A Study in Trust. Filosofiske Institutionen. Abo, 1996

[14] S. Marsh. Formalizing Trust as a Computational Concept. PhD thesis, University of Stirling, Scotland, 1994

[15] D. Romano. The Nature of Trust: Conceptual and Operational Clarification, Louisiana State University. PhD Thesis, 2003

[16] C. Ziegler, J. Golbeck. Investigating Correlations of Trust and Interest Similarity, Decision Support Systems 43(2), 460–475 (2007)

[17] EU COST Action IC0801 www,agreement-technologies.eu/wg5

[18] P. Matt, F. Toni. Combining Statistics and Arguments to compute Trust. AAMAS 2010, ACM press, Budapest.

[19] Walton, Douglas N. Argumentation schemes for presumptive reasoning. Routledge, 1996.

[20] Prakken, Henry. "On the nature of argument schemes." Dialectics, dialogue and argumentation. An examination of douglas Walton's theories of reasoning and argument (2010): 167-185.

[21] Lee, C. "Fuzzy logic in control systems: fuzzy logic controller" Systems, Man and Cybernetics, IEEE Transactions on 20.2 (1990): 404-418

[22] Parsons, Simon, et al. "Argument schemes for reasoning about trust."Computational Comma 2012 245 (2012): 430.

[23] Dondio, Pierpaolo et al. "Presumptive selection of trust evidence." Proceedings of AAMAS 2007. ACM, 2007

[24] Villata, S., et al. "A socio-cognitive model of trust using argumentation theory." Journal of Approximate Reasoning 54.4 (2013): 541-559.

[25] Dondio, Pierpaolo, and Stephen Barrett. "Comparison of six aggregation strategies to compute users' trustworthiness." Proceedings of the 19th ACM international conference on Information and knowledge management. ACM, 2010.